ABSTRACT

The present invention relates to a program to be used as a bioinformatics tool for analyzing files of nucleotide sequence data and finding the translation initiation codon. The invention described here uses Quadratic Discriminant Analysis (QDA) to determine the translation initiation codon in a nucleotide sequence. This program uses several components to provide a probability score for each potential translation initiation codon in a cDNA sequence. The components for each potential initiation codon in the sequence include a scoring method for (1) a model of the initiation consensus sequence, (2) the length of the upstream 5' untranslated region, (3) the third base composition downstream from the +6 position, (4) a codon transition score, and (5) a bulk monomer composition transition.